

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

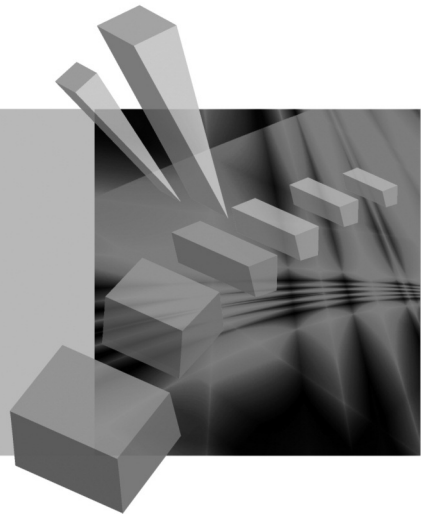
RESEARCH PAPERS

of Wrocław University of Economics

242

Taksonomia 19.

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2012

Recenzenci: Eugeniusz Gatnar, Elżbieta Gołata, Tadeusz Kufel, Józef Pocięcha,
Mirosław Szreder, Feliks Wysocki

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Tytuł sfinansowano ze środków Sekcji Klasyfikacji i Analizy Danych PTS
i Uniwersytetu Ekonomicznego we Wrocławiu

Publikacja jest dostępna na stronie www.ibuk.pl

Streszczenia opublikowanych artykułów są dostępne w międzynarodowej bazie danych
The Central European Journal of Social Sciences and Humanities <http://cejsh.icm.edu.pl>
oraz w The Central and Eastern European Online Library www.ceeol.com,
a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon [http://kangur.uek.krakow.pl/
bazy_ae/bazekon/nowy/index.php](http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php)

Informacje o naborze artykułów i zasadach recenzowania znajdują się
na stronie internetowej Wydawnictwa
www.wydawnictwo.ue.wroc.pl

Kopowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2012

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM
Nakład: 320 egz.

Spis treści

Wstęp	13
Stanisława Bartosiewicz , Jeszcze raz o skutkach subiektywizmu w analizie wielowymiarowej	17
Andrzej Sokolowski , Q uniwersalna miara odległości	22
Eugeniusz Gatnar , Jakość danych w systemach statystycznych banków centralnych (na przykładzie NBP)	31
Marek Walesiak , Pomiar odległości obiektów opisanych zmiennymi mierzonymi na skali porządkowej – strategię postępowania.....	39
Krzysztof Jajuga, Marek Walesiak , XXV lat konferencji taksonomicznych – fakty i refleksje	47
Józef Pocięcha, Barbara Pawelek , Model SEM w analizie zagrożenia bankructwem przedsiębiorstw w świetle koniunktury gospodarczej – problemy teoretyczne i praktyczne	50
Paweł Lula , Uczące się systemy pozyskiwania informacji z dokumentów tekstowych	58
Ewa Roszkowska , Zastosowanie metody TOPSIS do wspomaganie procesu negocjacji.....	68
Andrzej Młodak , Sąsiedztwo obszarów przestrzennych w ujęciu fizycznym oraz społeczno-ekonomicznym – podejście taksonomiczne	76
Andrzej Bąk , Modele kategorii nieuporządkowanych w badaniach preferencji	86
Jacek Kowalewski , Zintegrowany model optymalizacji badań statystycznych.....	96
Jan Paradysz, Karolina Paradysz , Obszary bezrobocia w Polsce – problem benchmarkowy.....	106
Tomasz Szubert , W co grać, aby jak najmniej przegrać? Próba klasyfikacji systemów gry w zakładach bukmacherskich.....	116
Izabela Szamrej-Baran , Klasyfikacja krajów UE ze względu na ubóstwo energetyczne	126
Sylwia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski , Analiza dojazdów do pracy za pomocą modelu grawitacji.....	135
Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz , Minimum egzystencji a czynniki warunkujące skłonność do korzystania z pomocy społecznej. Klasyfikacja gospodarstw domowych	144
Hanna Dudek , Subiektywne skale ekwiwalentności – analiza na podstawie danych o satysfakcji z osiągniętych dochodów	153

Joanicjusz Nazarko, Ewa Chodakowska, Marta Jaročka , Segmentacja szkół wyższych metodą analizy skupień <i>versus</i> konkurencja technologiczna ustalona metodą DEA – studium komparatywne.....	163
Ewa Chodakowska , Wybrane metody klasyfikacji w konstrukcji ratingu szkół.....	173
Bartosz Soliński , Sektor energetyki odnawialnej w krajach Unii Europejskiej – klasyfikacja w świetle strategii zarządzania zmianą.....	182
Krzysztof Szwarz , Klasyfikacja powiatów województwa wielkopolskiego ze względu na sytuację demograficzną.....	192
Elżbieta Gołata, Grażyna Dehnel , Rejestry administracyjne w analizie przedsiębiorczości.....	202
Katarzyna Chudy, Marek Sobolewski, Kinga Stępień , Wykorzystanie metod taksonomicznych w prognozowaniu wskaźników rentowności banków giełdowych w Polsce.....	212
Katarzyna Dębowska , Modelowanie upadłości przedsiębiorstw przy wykorzystaniu metod dyskryminacji i regresji.....	222
Alina Bojan , Wykorzystanie metod wielowymiarowej analizy danych do identyfikacji zmiennych wpływających na atrakcyjność wybranych inwestycji.....	231
Justyna Brzezińska , Analiza logarytmiczno-liniowa w badaniu przyczyn umieralności w krajach UE.....	240
Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka , Analiza klas ukrytych w badaniach satysfakcji studentów.....	247
Bartłomiej Jefmański , Pomiar opinii respondentów z wykorzystaniem elementów teorii zbiorów rozmytych i środowiska R.....	256
Julita Stańczuk , Porównanie rezultatów wielostanowej klasyfikacji obiektów ekonomicznych z wykorzystaniem analizy dyskryminacyjnej oraz sieci neuronowych.....	265
Jerzy Krawczuk , Skuteczność metod klasyfikacji w prognozowaniu kierunku zmian indeksu giełdowego S&P500.....	275
Anna Czapkiewicz, Beata Basiura , Symulacyjne badanie wpływu zaburzeń na grupowanie szeregów czasowych na podstawie modelu Copula-GARCH.....	283
Radosław Pietrzyk , Ocena efektywności inwestycji funduszy inwestycyjnych z tytułu doboru papierów wartościowych i umiejętności wykorzystania trendów rynkowych.....	291
Aleksandra Witkowska, Marek Witkowski , Zastosowanie metody Panzara-Rosse’a do pomiaru poziomu konkurencji w sektorze banków spółdzielczych.....	306
Marcin Pelka , Podejście wielomodelowe z wykorzystaniem metody <i>boosting</i> w analizie danych symbolicznych.....	315
Justyna Wilk , Analiza porównawcza oprogramowania komputerowego w klasyfikacji danych symbolicznych.....	323

Tomasz Bartłomowicz, Justyna Wilk , Zastosowanie metod analizy danych symbolicznych w przeszukiwaniu dziedzinowych baz danych.....	333
Kamila Migdał-Najman , Propozycja hybrydowej metody grupowania opartej na sieciach samouczących	342
Dorota Rozmus , Porównanie dokładności taksonomii spektralnej oraz zagregowanych algorytmów taksonomicznych opartych na idei metody <i>bagging</i>	352
Krzysztof Najman , Grupowanie dynamiczne z wykorzystaniem samouczących się sieci GNG	361
Małgorzata Misztal , Wpływ wybranych metod uzupełniania brakujących danych na wyniki klasyfikacji obiektów z wykorzystaniem drzew klasyfikacyjnych w przypadku zbiorów danych o niewielkiej liczebności – ocena symulacyjna	370
Mariusz Kubus , Zastosowanie wstępnego uwarunkowania zmiennej objaśnianej do selekcji zmiennych.....	380
Barbara Batóg, Jacek Batóg , Wykorzystanie analizy dyskryminacyjnej do identyfikacji czynników determinujących stopę zwrotu z inwestycji na rynku kapitałowym	387
Katarzyna Wójcik, Janusz Tuchowski , Analiza porównawcza miar podobieństwa tekstów opartych na macierzy częstości i tekstów opartych na wiedzy dziedzinowej	396
Iwona Staniec , Analiza czynnikowa w identyfikacji obszarów determinujących doskonalenie systemów zarządzania w polskich organizacjach	406
Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawełczyk, Jerzy Kołodziej, Jerzy Błaszczyk , Analiza porównawcza wybranych technik eksploracji danych do klasyfikacji danych medycznych z brakującymi obserwacjami	416
Iwona Foryś , Wykorzystanie analizy log-liniowej do wyboru czynników determinujących atrakcyjność cenową mieszkań w obrocie wtórnym na przykładzie lokalnego rynku mieszkaniowego.....	426
Ewa Genge , Analiza skupień oparta na mieszankach uciętych rozkładów normalnych.....	436
Jerzy Korzeniewski , Ocena efektywności metody uśredniania zmiennych i metody Ichino selekcji zmiennych w analizie skupień	444
Andrzej Dudek , SMS – propozycja nowego algorytmu analizy skupień	451
Artur Mikulec , Metody oceny wyniku grupowania w analizie skupień.....	460
Małgorzata Machowska-Szewczyk , Algorytm klasyfikacji rozmytej dla obiektów opisanych za pomocą zmiennych symbolicznych oraz rozmytych	469
Artur Zaborski , Analiza PROFIT i jej wykorzystanie w badaniu preferencji	479
Karolina Bartos , Analiza skupień wybranych państw ze względu na strukturę wydatków konsumpcyjnych obywateli – zastosowanie sieci Kohonena	488

Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , Klasyfikacja gospodarstw domowych ze względu na bodźce do zawierania umowy o ubezpieczenie z wykorzystaniem modeli zmiennych jakościowych .	496
Izabela Kurzawa , Zastosowanie modelu LA/AIDS do badania elastyczności cenowych popytu konsumpcyjnego w gospodarstwach domowych w relacji miasto–wieś	505
Aleksandra Łuczak, Feliks Wysocki , Metody porządkowania liniowego obiektów opisanych za pomocą cech metrycznych i porządkowych	513
Agnieszka Sompolska-Rzechuła , Porównanie klasycznej i pozycyjnej taksonomicznej analizy zróżnicowania jakości życia w województwie zachodniopomorskim	523
Joanna Banaś, Małgorzata Machowska-Szewczyk , Ocena intensywności wykorzystania skrzynek poczty elektronicznej za pomocą uporządkowanego modelu probitowego	532
Iwona Bąk , Segmentacja gospodarstw domowych emerytów i rencistów pod względem wydatków na rekreację i kulturę	541
Aneta Becker , Zastosowanie metody ANP do porządkowania województw Polski pod względem dynamiki wykorzystania ICT w latach 2008-2010	552
Katarzyna Dębowska , Klasyfikacja sektorów ze względu na ich kondycję finansową przy użyciu metod wielowymiarowej analizy statystycznej	562
Anna Domagała , Propozycja metody doboru zmiennych do modeli DEA (procedura kombinowanego doboru w przód).....	571
Henryk Gierszal, Karina Pawlina, Maria Urbańska , Analiza statystyczna w badaniach zapotrzebowania na usługi teleinformatyczne sieci łączności ruchomej	580
Hanna Gruchociak , Konstrukcja estymatora regresyjnego dla danych o strukturze dwupoziomowej.....	590
Tomasz Klimanek, Marcin Szymkowiak , Zastosowanie estymacji pośredniej uwzględniającej korelację przestrzenną w opisie niektórych charakterystyk rynku pracy	601
Jarosław Lira , Prognozowanie opłacalności produkcji żywca wieprzowego w Polsce	610
Christian Lis , Wykorzystanie metody klasyfikacji w ocenie konkurencyjności portów południowego Bałtyku	619
Beata Bieszk-Stolorz, Iwona Markowicz , Wykorzystanie wielomianowego modelu logitowego do oceny szansy podjęcia pracy przez bezrobotnych .	628
Lucyna Przezbórska-Skobiej, Jarosław Lira , Przestrzeń agroturystyczna Polski i ocena jej atrakcyjności.....	637
Paweł Ulman , Model rozkładu wydatków a funkcje popytu.....	646
Maria Urbańska, Tadeusz Mizera, Henryk Gierszal , Zastosowanie metod analizy statystycznej w badaniach mięczaków	655

Summaries

Stanisława Bartosiewicz , The effects of subjectivism in multivariate analysis revisited.....	21
Andrzej Sokółowski , Q universal distance measure	30
Eugeniusz Gatnar , Data quality in central banks' statistical systems (NBP example)	38
Marek Walesiak , Distance measures for ordinal data – strategies of proceedings.....	46
Krzysztof Jajuga, Marek Walesiak , XXV years of taxonomic conferences – some facts and remarks.....	49
Józef Pocięcha, Barbara Pawelek , General SEM model in researching corporate bankruptcy and business cycles – theoretical and practical problems.....	57
Paweł Lula , Learning-based systems of information extraction from textual resources	67
Ewa Roszkowska , The application of the TOPSIS method to support the negotiation process	75
Andrzej Młodak , Neighborhood of spatial areas in the physical and socio-economic context – a taxonomic approach.....	85
Andrzej Bąk , Models for unordered categories in preference analysis.....	95
Kowalewski Jacek , An integrated model of optimizing statistical surveys	105
Jan Paradysz, Karolina Paradysz , Areas of unemployment in Poland – benchmark problem	115
Tomasz Szubert , How to play to lose the least? Classification of systems in sports bets	125
Izabela Szamrej-Baran , Classification of EU member states in view of fuel poverty	134
Sylvia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski , An attempt to use the gravity model in the analysis of commuters.....	143
Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz , Subsistence minimum versus factors influencing tendency to benefit from social care. Classification of households	152
Hanna Dudek , Subjective equivalence scales – analysis based on data about satisfaction with incomes.....	162
Joanicjusz Nazarko, Ewa Chodakowska, Marta Jarocka , Segmentation of universities using cluster analysis versus technological competitors determined by the DEA method – a comparative study	172
Ewa Chodakowska , Selected methods of classification in schools' rating.....	181
Bartosz Soliński , Renewable energy sector in the European Union – classification in the light of change management strategy	191
Krzysztof Szwarz , Classification of Wielkopolska voivodeship due to the demographic situation	201

Elżbieta Gołata, Grażyna Dehnel , Administrative registers in business analysis.....	211
Katarzyna Chudy, Marek Sobolewski, Kinga Stępień , Application of taxonomic methods in forecasting the profitability ratios of listed banks in Poland.....	221
Katarzyna Dębowska , Modeling bankruptcy of firms by using discrimination and regression methods.....	230
Alina Bojan , Identification of variables which influence attractiveness of given investments with the usage of multivariate analysis.....	239
Justyna Brzezińska , Log-linear analysis in the study of mortality in EU.....	246
Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka , Latent class analysis in student satisfaction surveys.....	254
Bartłomiej Jefmański , The respondent's opinions measurement in the R program with an application of fuzzy sets theory.....	264
Julita Stańczuk , A comparison of the results of multistate classification of economic objects using discriminant analysis and artificial neural networks.....	274
Jerzy Krawczuk , Effectiveness of classification methods in S&P500 stock index direction changes forecasting.....	282
Anna Czapkiewicz, Beata Basiura , The simulation study of the utility of the Copula-GARCH models for clustering financial time series.....	290
Radosław Pietrzyk , Timing and selectivity in mutual funds performance measurement.....	305
Aleksandra Witkowska, Marek Witkowski , Use of the Panzar-Rosse method to assess of the competition level in the cooperative banks sector.....	314
Marcin Pelka , Ensemble learning with the application of <i>boosting</i> in symbolic data analysis.....	322
Justyna Wilk , Comparative study of symbolic data classification software.....	332
Tomasz Bartłomowicz, Justyna Wilk , Application of symbolic data analysis methods for domain database searching.....	341
Kamila Migdał-Najman , A proposal of hybrid clustering method based on self-learning networks.....	351
Dorota Rozmus , Comparison of accuracy of spectral clustering and cluster ensembles stability based on bagging idea.....	360
Krzysztof Najman , A dynamic grouping based on self-learning GNG networks.....	369
Małgorzata Misztal , Influence of data imputation methods on the results of object classification using classification trees in the case of small data sets – simulation assessment.....	379
Mariusz Kubus , The application of pre-conditioning of explanatory variable for feature selection.....	386
Barbara Batóg, Jacek Batóg , Application of discriminant analysis to the identification of factors determining the rate of return on the capital market.....	395

Katarzyna Wójcik, Janusz Tuchowski , Comparative analysis of text documents similarity measures based on frequency matrix and based on domain knowledge.....	405
Iwona Staniec , Factor analysis in the identification of areas that determine the improvement of management systems in Polish organizations.....	415
Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawełczyk, Jerzy Kołodziej, Jerzy Błaszczyk , Comparative analysis of selected data mining approaches to the classification of medical data with missing values (covariates).....	425
Iwona Foryś , The log-linear analysis using to select the factors determining the attractiveness of the price of flats on the secondary market on the example of local housing market.....	435
Ewa Genge , Trimming approach to the mixtures of normal distributions.....	443
Jerzy Korzeniewski , Efficiency assessment of Ichino method and mean value method of selecting variables in cluster analysis.....	450
Andrzej Dudek , SMS – proposal of new clustering algorithm.....	459
Artur Mikulec , Evaluation methods for the grouping result in cluster analysis.....	468
Małgorzata Machowska-Szewczyk , Fuzzy clustering algorithm for objects described by symbolic or fuzzy variables.....	478
Artur Zaborski , PROFIT analysis and its using in the research of preferences.....	487
Karolina Bartos , Cluster analysis of selected countries due to the structure of their citizens' consumer expenditures – the use of Kohonen networks.....	495
Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , Classification of households according to the impulses of concluding the insurance contract by means of qualitative variable models.....	504
Izabela Kurzawa , The application of LA/AIDS model to examine price elasticities of demand of households in the urban-rural relationship.....	512
Aleksandra Luczak, Feliks Wysocki , Linear ordering methods of objects described by a set of metric and ordinal characteristics.....	522
Agnieszka Sompolska-Rzechuła , The comparison of the classical and positional taxonomic analysis of the quality of life differentiation in Zachodniopomorskie voivodeship.....	531
Joanna Banaś, Małgorzata Machowska-Szewczyk , Evaluation of intensity of mailboxes using with the ordered probit model.....	540
Iwona Bąk , Segmentation of pensioners and annuitants households in terms of expenditures on recreation and culture.....	551
Aneta Becker , Application of ANP method to organize Polish voivodships in terms of dynamics of the use of ICT in 2008-2010.....	561
Katarzyna Dębowska , The classification of sectors' financial situation using the methods of multivariate statistical analysis.....	570

Anna Domagała , Proposal of a new method for variable selection in DEA models (combined forward stepwise selection method).....	579
Henryk Gierszal, Karina Pawlina, Maria Urbańska , Statistical analysis in demand research of ICT services in mobile networks.....	589
Hanna Gruchociak , Construction of regression estimator for two-level data	600
Tomasz Klimanek, Marcin Szymkowiak , Application of spatial models in indirect estimation of some labor market characteristics	609
Jarosław Lira , Forecasting of hog livestock production profitability in Poland	618
Christian Lis , The utilization of taxonomic methods in the appraisal of competitiveness of south Baltic ports	627
Beata Bieszk-Stolorz, Iwona Markowicz , The application of the multinomial logit model in evaluating employment odds for the unemployed job seekers	636
Lucyna Przezbórska-Skobiej, Jarosław Lira , Agritourism space of Poland and its valuation.....	645
Paweł Ulman , Model of expenses distribution and demand functions.....	654
Maria Urbańska, Tadeusz Mizera, Henryk Gierszal , Methods of statistical analysis in research of molluscs	663

Bartłomiej Jefmański

Uniwersytet Ekonomiczny we Wrocławiu

POMIAR OPINII RESPONDENTÓW Z WYKORZYSTANIEM ELEMENTÓW TEORII ZBIORÓW ROZMYTYCH I ŚRODOWISKA R

Streszczenie: Opracowanie stanowi propozycję pomiaru opinii respondentów polegającą na połączeniu zastosowania dyskretnych skal szacunkowych oraz liczb rozmytych. Zaprezentowano dwa warianty konstrukcji tzw. pytań pomocniczych w kwestionariuszu ankiety, które umożliwiają przekształcenie wyrażań lingwistycznych w liczby rozmyte (np. o trójkątnych lub trapezoidalnych funkcjach przynależności). Proponowane podejście zastosowano w badaniu opinii doradców zawodowych na temat jakości kształcenia ustawicznego w podregionie wałbrzyskim. Wyniki analizy potwierdziły różnice w interpretacji przez respondentów wartości lingwistycznych stanowiących punkty na zastosowanej w opracowaniu skali szacunkowej. Operacje arytmetyczne na liczbach rozmytych niezbędne do analizy danych ankietowych z zastosowaniem proponowanego podejścia przeprowadzono w programie **R**.

Słowa kluczowe: liczby rozmyte, wartości lingwistyczne, dyskretne skale szacunkowe.

1. Wstęp

Pomiar opinii i postaw respondentów odbywa się zazwyczaj za pomocą skal szacunkowych, które z punktu widzenia teorii pomiaru należą do grupy skal porządkowych. W badaniach marketingowych najczęściej wśród skal szacunkowych spotyka się skale graficzne dyskretne, czyli takie, które określają zarówno skrajne, jak i pośrednie jej wartości (punkty skali) [Sagan 2004, s. 79-80]. Klasyczne podejście zwiększające użyteczność tego typu skal zakłada przyporządkowanie poszczególnym ich punktom tzw. wartości lingwistycznych. Przykładami stosowanych wartości lingwistycznych mogą być określenia typu „dobrze”, „nisko”, „często” itp. Taki sposób pomiaru umożliwia wyrażenie opinii przez respondentów w naturalnym języku, dlatego jest atrakcyjny z punktu widzenia respondenta, ale stanowi jednocześnie wyzwanie dla badacza, który winien uwzględnić nieprecyzyjność i niejednoznaczność stosowanych wyrażań. Jednym z nowatorskich podejść w tym zakresie jest zastosowanie liczb rozmytych. Rozwiązanie to polega na przyporządkowaniu wartościom lingwistycznym przypisanym poszczególnym punktom skali pomiaru liczb rozmytych. Niniejsze opracowanie stanowi jedną z propozycji w tym zakresie polegającą

na umieszczeniu w kwestionariuszu ankiety stosownych pytań, które umożliwiają na etapie analizy wyników liczbową interpretację wartości lingwistycznych. Proponowane podejście scharakteryzowano na podstawie przeprowadzonego przez autora opracowania badania opinii doradców zawodowych na temat jakości kształcenia ustawicznego w podregionie wałbrzyskim.

2. Charakterystyka analizy wyników pomiaru na skalach szacunkowych z zastosowaniem liczb rozmytych i środowiska \mathbf{R}

Liczba rozmyta to zbiór rozmyty $\tilde{A} \subseteq R$ określony w zbiorze liczb rzeczywistych spełniający następujące warunki [Zimmermann 2001, s. 59]:

- \tilde{A} jest zbiorem normalnym,
- \tilde{A} jest zbiorem wypukłym,
- funkcja przynależności zbioru \tilde{A} jest funkcją kawałkami ciągłą.

Zastosowanie liczb rozmytych na etapie pomiaru opinii respondentów za pomocą skal szacunkowych sprowadza się do przyporządkowania poszczególnym wartościom lingwistycznym, stanowiącym punkty skali, liczb rozmytych o stosowanej funkcji przynależności i zakresie jej dziedziny. Wybrane z literatury przedmiotu przykłady przyporządkowania liczb rozmytych wartościom lingwistycznym wraz z charakterystyką funkcji przynależności i zakresami dziedzin liczb rozmytych wyszczególniono m.in. w opracowaniu Jefmańskiego [2011, s. 294-298].

Statystyczna analiza uzyskanych w ten sposób wyników pomiaru wymaga znajomości arytmetyki liczb rozmytych. Pomocnym narzędziem w tym zakresie może być program \mathbf{R} , w którym analiza pomiaru opinii respondentów z zastosowaniem liczb rozmytych możliwa jest w ramach dwóch pakietów: *SAFD* oraz *fuzzyOP*.

Pakiet *SAFD* oferuje rozbudowane funkcje do statystycznej analizy danych rozmytych. Oprócz podstawowych działań arytmetycznych na liczbach rozmytych pozwala na obliczanie takich statystyk, jak wariancja, kowariancja czy też korelacja dla danych rozmytych. Ma wbudowaną funkcję losowego generowania liczb rozmytych oraz umożliwia szacowanie funkcji regresji dla liczb rozmytych o trapezoidalnej postaci funkcji przynależności. Ma również funkcję pozwalającą na wyostrzenie (defuzyfikację) wyników analiz otrzymanych w postaci liczb rozmytych.

Pakiet *fuzzyOP* w całości został poświęcony wyłącznie operacjom arytmetycznym na liczbach rozmytych. Zaimplementowane funkcje umożliwiają przeprowadzenie większości operacji arytmetycznych na liczbach rozmytych oraz pozwalają na wizualizację otrzymanych wyników. W pakiecie każda liczba rozmyta jest opisana za pomocą dwóch wektorów, przy czym elementy pierwszego z nich określają zakresy dziedzin liczb rozmytych, a elementy drugiego wektora wskazują na odpowiadające im stopnie przynależności do zbioru rozmytego [Aklan i in. 2009, s. 2]. I tak np. obliczenie średniej arytmetycznej trzech liczb rozmytych o odpowiednio następujących zakresach dziedziny: $\tilde{A} = (0; 0; 0, 3)$, $\tilde{B} = (0; 0, 25; 0, 5)$, $\tilde{C} = (0, 5; 0, 75; 1)$ możliwe jest przez zastosowanie następującej składni poleceń:

```

> library(fuzzyOP)
> A1<-c(0,0,0.3)
> a1<-c(0,1,0)
> B2<-c(0,0.25,0.5)
> b2<-c(0,1,0)
> C3<-c(0.5,0.75,1)
> c3<-c(0,1,0)
> A<-cbind(A1,a1,B2,b2,C3,c3)
> srednia<-fuzzymeans(100, A, c(1,2,3),2)
> srednia.

```

Bardziej szczegółową charakterystykę składni wybranych funkcji wraz z poleceniami umożliwiającymi generowanie wyników operacji arytmetycznych na liczbach rozmytych oraz wizualizację wyników zaprezentowano m.in. w opracowaniu Jefmańskiego [2012].

3. Analiza wyników badania ankietowego z zastosowaniem liczb rozmytych

Proponowane w artykule połączenie pomiaru na skali szacunkowej z liczbami rozmytymi przyporządkowanymi poszczególnym wartościom skali zastosowano w badaniu ankietowym przeprowadzonym w okresie 01.08.2011-30.08.2011 na terenie podregionu wałbrzyskiego (powiat wałbrzyski, świdnicki, dzierzoniowski, ząbkowicki, kłodzki). Celem badania było poznanie opinii doradców zawodowych nt. wybranych aspektów jakości kształcenia ustawicznego w podregionie wałbrzyskim. Próbę badawczą stanowiło 23 doradców zawodowych zatrudnionych w powiatowych urzędach pracy, centrach informacji zawodowej, ochotniczych hufcach pracy, dobranych w sposób celowy. Do udziału w badaniu zaproszono wyłącznie doradców zawodowych, którzy mają dużą wiedzę na temat funkcjonowania lokalnego rynku pracy, zdobytą m.in. przez udział w projektach badawczych prowadzonych na terenie podregionu (niektórzy z respondentów biorą również czynny udział w zawiązanych na terenie podregionu partnerstwach lokalnych mających na celu m.in. prognozowanie zmian na lokalnym rynku pracy).

Na potrzeby niniejszego opracowania analizie poddano odpowiedzi na jedno z pytań dotyczące oceny wybranych czynników związanych z rozwojem kształcenia ustawicznego na terenie podregionu wałbrzyskiego w brzmieniu: „Jak ocenia Pan(i) niżej wyszczególnione czynniki związane z rozwojem kształcenia ustawicznego na terenie podregionu wałbrzyskiego”. Zastosowano pięciostopniową szacunkową skalę pomiaru o następujących kategoriach: 1 – „bardzo nisko”, 2 – „nisko”, 3 – „średnio”, 4 – „wysoko”, 5 – „bardzo wysoko”. Respondenci oceniali następujące czynniki:

X_1 – chęć uczenia się mieszkańców podregionu wałbrzyskiego,

- X_2 – dostępność aktualnej informacji o rynku pracy i potrzebach szkoleniowych mieszkańców i przedsiębiorców,
- X_3 – funkcjonowanie systemu zewnętrznego mierzenia jakości i akredytacji,
- X_4 – koordynacja działań w zakresie kształcenia ustawicznego,
- X_5 – promocja kształcenia ustawicznego przez władze samorządowe,
- X_6 – upowszechnianie pozytywnych przykładów wpływu szkoleń na funkcjonowanie przedsiębiorstw,
- X_7 – sposób i zakres funkcjonowania systemu zewnętrznemu uznawanych kwalifikacji,
- X_8 – zaangażowanie nauczycieli w promowanie aktywności edukacyjnej wśród dzieci i młodzieży,
- X_9 – chęć komunikacji, współpracy i wymiany informacji między instytucjami zaangażowanymi w kształcenie ustawiczne,
- X_{10} – usługi doradztwa w zakresie planowania kariery zawodowej.

Rzetelność zastosowanej skali pomiaru oceniono za pomocą statystyki α -Cronbacha, która wyniosła 0,84. Wartość średniej korelacji między pozycjami skali ukształtowała się na poziomie 0,37. Ze względu na wysoką wartość statystyki α -Cronbacha do dalszych analiz wykorzystano pełny zbiór zmiennych.

W kwestionariuszu ankiety zamieszczono dwa warianty pytania pomocniczego, którego wyniki pozwoliły przyporządkować liczby rozmyte poszczególnym wartościom lingwistycznym (zob. tab. 1-2).

Tabela 1. Pytanie pomocnicze – wariant nr 1

Jaki odsetek osób bezrobotnych, które znalazły zatrudnienie dzięki skorzystaniu z usług doradców zawodowych, uznałby/uznałaby Pan/i za:

Podpowiedź: dla każdej z pięciu kategorii (bardzo niski, niski itd.) wyszczególnionych w pierwszej kolumnie poniższej tabeli proszę przyporządkować trzy liczby z zakresu [0-100%]. Np. kategoria „średni” – dolna granica 40%, wartość optymalna 50%, górna granica 55%.

Kategoria	Dolna granica	Wartość optymalna	Górna granica
Bardzo niski	0 % % %
Niski % % %
Średni % % %
Wysoki % % %
Bardzo wysoki % %	100 %

Tabela 2. Pytanie pomocnicze – wariant nr 2

Jaki odsetek absolwentów ponadgimnazjalnych szkół zawodowych, którzy znaleźli zatrudnienie bezpośrednio po ukończeniu szkoły, uznałby/uznałaby Pan/i za:

Podpowiedź: dla każdej z pięciu kategorii (bardzo niski, niski itd.) proszę przyporządkować liczbę z zakresu [0-100%]. Np. „niski – około 20% osób”.

Bardzo niski	około % osób
Niski	około % osób
Średni	około % osób
Wysoki	około % osób
Bardzo wysoki	około % osób

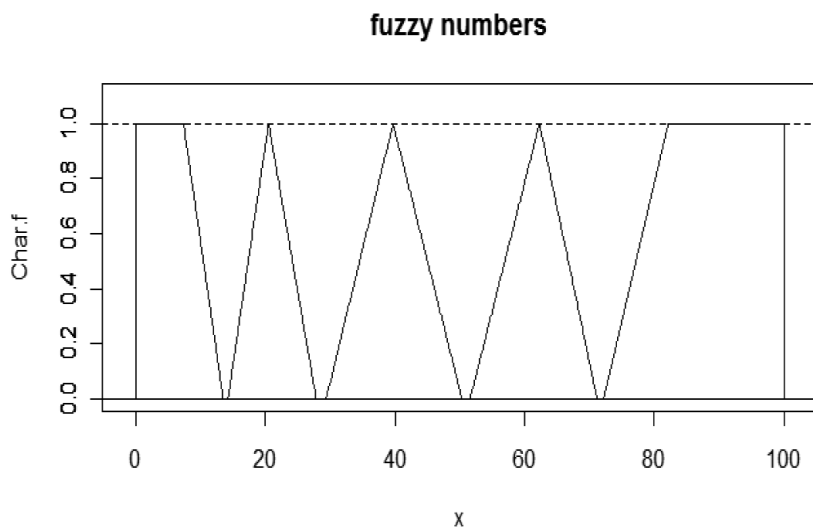
Źródło: opracowanie własne.

W pierwszym wariantcie każdy z respondentów dla każdej z pięciu wartości lingwistycznych przyporządkował trzy liczby z zakresu [0-100%] – wartość najlepiej oddającą jego zdaniem znaczenie danej wartości oraz ewentualnie dopuszczalną dolną i górną granicę odchyień. Tak sformułowane pytanie pozwala w bezpośredni sposób oszacować dla każdego respondenta znaczenie poszczególnych kategorii przez przyporządkowanie im liczb rozmytych, których środkami są wartości optymalne, a lewymi i prawymi zakresami dziedzin odpowiednio dolne i górne granice. Tak otrzymane wyniki można również uśrednić dla całej próby badawczej przez obliczenie średnich liczb rozmytych dla każdej z pięciu kategorii.

Drugi wariant pytania pomocniczego przewidywał przyporządkowanie każdej z pięciu wartości lingwistycznych tylko jednej wartości liczbowej z przedziału [0-100%]. Takie podejście różni się od poprzedniego m.in. tym, że nie dysponujemy dla każdego z respondentów indywidualnymi ocenami poszczególnych wartości lingwistycznych w postaci liczb rozmytych. Tak sformułowane pytanie umożliwia przyporządkowanie liczb rozmytych poszczególnym wartościom skali tylko na poziomie całej próby badawczej. Proponuje się ustalenie dla każdej z wartości lingwistycznych wartości dominującej, która będzie stanowiła środek liczby rozmytej. Lewy i prawy zakres dziedziny mogą stanowić odpowiednio minimalne i maksymalne oceny nadane przez respondentów poszczególnym wartościom lingwistycznym. Otrzymamy w ten sposób pięć liczb rozmytych dla każdej z pięciu wartości zastosowanej skali szacunkowej.

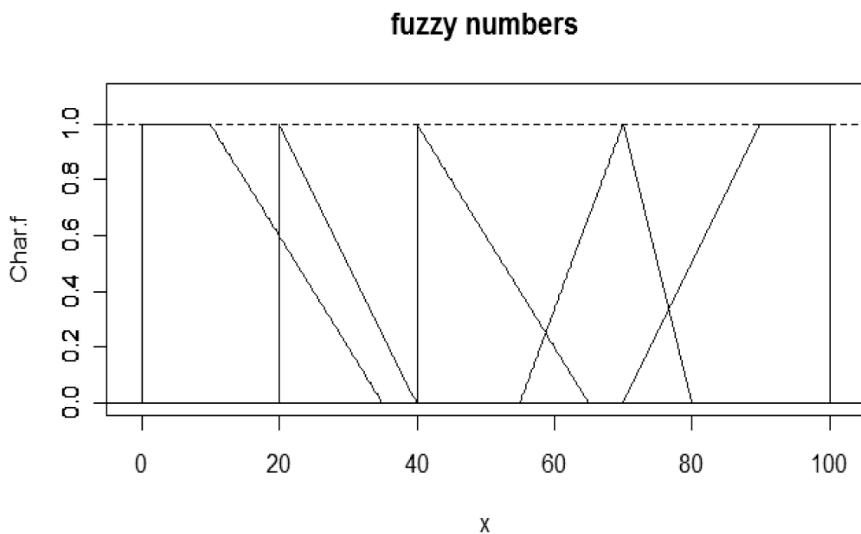
Kształt i rozpiętość liczb rozmytych przyporządkowanych wartościom lingwistycznym zastosowanym w badaniu dla dwóch wariantów pytania pomocniczego zaprezentowano na rys. 1-2.

Zakresy dziedzin i środki liczb rozmytych dla poszczególnych wartości lingwistycznych wyszczególniono w tab. 3.



Rys. 1. Uśrednione dla całej próby badawczej liczby rozmyte przyporządkowane poszczególnym wartościom lingwistycznym (dla wariantu pierwszego)

Źródło: opracowanie własne.



Rys. 2. Liczby rozmyte przyporządkowane poszczególnym wartościom lingwistycznym na poziomie całej próby badawczej (dla wariantu drugiego)

Źródło: opracowanie własne.

Tabela 3. Zakresy dziedzin i środki liczb rozmytych dla poszczególnych wartości lingwistycznych

Kategoria (wartość lingwistyczna)	Pierwszy wariant			Drugi wariant		
	lewy zakres dziedziny	środek liczby rozmytej*	prawy zakres dziedziny	lewy zakres dziedziny	środek liczby rozmytej*	prawy zakres dziedziny
Bardzo nisko	0%	[0%-7,42%]	13,53%	0%	[0%-10%]	35%
Nisko	14,26%	20,53%	27,89%	20%	20%	40%
Średnio	29,42%	39,74%	50,37%	40%	40%	65%
Wysoko	51,63%	62,32%	71,21%	55%	70%	80%
Bardzo wysoko	72,21%	[82,16%-100%]	100%	70%	[90%-100%]	100%

* Kategoriom skrajnym „bardzo nisko” i „bardzo wysoko” przyporządkowano trapezoidalne liczby rozmyte, dlatego funkcja przynależności osiąga wartość równą 1 w pewnym przedziale, a nie dla pojedynczej wartości, tak jak ma to miejsce w przypadku pozostałych kategorii, którym przyporządkowano trójkątne liczby rozmyte.

Źródło: opracowanie własne.

Jeśli chodzi o pierwszy wariant, charakterystyczną rzeczą, którą można zauważyć na rys. 1 jest to, że im bardziej na prawo skali znajduje się wartość lingwistyczna, tym większa rozpiętość zakresu dziedziny przyporządkowanej jej liczby rozmytej. Oznacza to, że respondenci uczestniczący w badaniu mają bardziej sprecyzowane pojęcia „bardzo nisko”, „nisko” niż pozostałe kategorie. Różnice można zaobserwować szczególnie w przypadku kategorii „bardzo wysoko”. Skrajnym punktem skali przyporządkowano liczby rozmyte o trapezoidalnej funkcji przynależności, co wynika ze sposobu sformułowania pytania pomocniczego. Charakterystyczne jest to, że zakresy dziedzin żadnej z liczb rozmytych przyporządkowanych do poszczególnych wartości skali nie pokrywają się.

W wariancie drugim, podobnie jak w pierwszym, skrajnym wartościom zostały przyporządkowane liczby rozmyte o trapezoidalnej funkcji przynależności. Ponadto w przypadku obu tych liczb zaobserwowano stosunkowo dużą rozpiętość liczb rozmytych. Jeśli chodzi o pozostałe wartości lingwistyczne, to można mówić o podobnych rozpiętościach dziedzin liczb rozmytych. Należy również zwrócić uwagę na liczby rozmyte przyporządkowane kategoriom „niski” i „średni” – ich lewe zakresy dziedzin są równe wartościom środkowym.

Dysponując oszacowanymi na poziomie całej próby badawczej liczbami rozmytymi dla każdej z pięciu wartości lingwistycznych stosowanych w kwestionariuszu ankiety, obliczono średnie oceny poszczególnych czynników związanych z rozwojem kształcenia ustawicznego w podregionie wałbrzyskim. Ponieważ średnie oceny miały postać liczb rozmytych, porównanie czynników pod względem ich ważności wymagało zastosowania jednej z metod defuzyfikacji (wyostrzania). Wyniki, po wyostrzeniu metodą środka ciężkości, scharakteryzowaną m.in. w opracowaniu Opricovica i Tzenga [2003, s. 641-642], zaprezentowano dla dwóch wariantów pytania pomocniczego (zob. tab. 4-5).

Tabela 4. Wyniki oceny czynników wpływających na rozwój kształcenia ustawicznego w podregionie wałbrzyskim

Czynnik	Ocena czynników	
	wariant nr 1	wariant nr 2
X_1	36,39	43,43
X_2	36,18	44,26
X_3	33,14	41,30
X_4	31,81	39,54
X_5	33,19	40,56
X_6	34,08	41,85
X_7	40,28	48,43
X_8	48,84	56,67
X_9	33,58	41,76
X_{10}	57,99	65,28

Źródło: obliczenia własne z zastosowaniem pakietu *fuzzyOP* w programie **R**.

Tabela 5. Ranking czynników wpływających na rozwój kształcenia ustawicznego w podregionie wałbrzyskim

Czynnik	Pozycja czynnika w rankingu	
	wariant nr 1	wariant nr 2
X_1	4	5
X_2	5	4
X_3	9	8
X_4	10	10
X_5	8	9
X_6	6	6
X_7	3	3
X_8	2	2
X_9	7	7
X_{10}	1	1

Źródło: obliczenia własne z zastosowaniem pakietu *fuzzyOP* w programie **R**.

Analizując wyniki z tab. 4-5, można stwierdzić, że pozycje czynników w rankingach ustalonych dla dwóch wariantów analizy są bardzo podobne. Czynniki, które zdaniem respondentów mają największy wpływ na rozwój kształcenia ustawicznego w podregionie wałbrzyskim, to:

- usługi doradztwa w zakresie planowania kariery zawodowej,
- zaangażowanie nauczycieli w promowanie aktywności edukacyjnej wśród dzieci i młodzieży,
- sposób i zakres funkcjonowania systemu zewnętrznemu uznawanych kwalifikacji.

Cztery spośród dziesięciu czynników (X_1, X_2, X_3, X_5) zajęły różne pozycje w dwóch rankingach, przy czym przesunięcia te dotyczyły w każdym z przypadków zmiany wyłącznie o jedną pozycję. W przypadku pozostałych czynników otrzymano jednakowe wyniki dla dwóch wariantów.

4. Podsumowanie

W opracowaniu scharakteryzowano propozycję pomiaru opinii respondentów polegającą na połączeniu zastosowania skal szacunkowych oraz liczb rozmytych. Zaprezentowano dwa sposoby konstrukcji tzw. pytań pomocniczych, które umożliwiają przekształcenie wyrażań lingwistycznych w liczby rozmyte, np. o trójkątnych i tra-

pezoidalnych funkcjach przynależności. Należy podkreślić, że w opinii respondentów zdecydowanie łatwiejszy i bardziej zrozumiały był wariant drugi. Warto również zaznaczyć, że otrzymane wyniki (szczególnie kształt liczb rozmytych) dla wariantu pierwszego sugerują, że respondenci w przypadku określania dolnych i górnych wartości dla poszczególnych wartości lingwistycznych w większości przypadków ustalają jednakowe rozpiętości lewych i prawych zakresów dziedzin. Ponadto rozpiętości te dla kategorii bezpośrednio ze sobą sąsiadujących w większości przypadków są przez respondentów tak ustalone, aby się nie pokrywały. Ze względu na fakt, że wyniki otrzymane dla dwóch wariantów analizy są bardzo zbliżone, obiecujący wydaje się sposób formułowania odpowiedzi na pytanie pomocnicze przedstawiony w drugim wariantcie.

Literatura

- Aklan S., Altindas E., Macit R., Umar S., Unal H., *Fuzzy numbers and the main mathematical operations*, 2009, <http://cran.r-project.org/web/packages/fuzzyOP/fuzzyOP.pdf>.
- Jefmański B., *Pomiar i ocena jakości usług z zastosowaniem liczb rozmytych – aspekty metodologiczne i przykłady zastosowań*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 151, Wydawnictwo UE, Wrocław 2011.
- Jefmański B., *Rozmyty pomiar i analiza jakości usług z zastosowaniem liczb rozmytych i środowiska R*, Zeszyty Naukowe Uniwersytetu Ekonomicznego w Poznaniu, Wydawnictwo UE, Poznań 2012 (w druku).
- Opricovic S., Tzeng G.-H., *Defuzzification within a multicriteria decision model*, International Journal of Uncertainty, „Fuzziness and Knowledge-Based Systems” 2003, vol. 11, no 5.
- Sagan A., *Badania marketingowe. Podstawowe kierunki*, AE, Kraków 2004.
- Trutschnig W., *Statistical analysis of fuzzy data*, <http://cran.r-project.org/web/Packages/SAFD/SAFD.Pdf>.
- Zimmermann H.-J., *Fuzzy Set Theory and Its Applications*, Kluwer Academic Publishers, Boston 2001.

THE RESPONDENTS' OPINIONS MEASUREMENT IN THE R PROGRAM WITH AN APPLICATION OF FUZZY SETS THEORY

Summary: The article is a proposition for the respondents' opinions measurement consisting in connecting an application of discrete rating scales and fuzzy numbers. The study introduces and characterizes two variants of special questions for transforming linguistic variables into fuzzy numbers (e.g. triangular and trapezoidal). The proposed approach was used in the opinion poll of some career advisors on the quality of the lifelong education in Wałbrzych subregion. The results of the analysis confirmed the differences in the interpretation by the respondents of the linguistic values constituting the points of the discrete rating scale applied in the article. The arithmetic operations on fuzzy numbers indispensable to analyze the survey data using the proposed approach were computed in the R program.

Keywords: fuzzy numbers, linguistic values, discrete rating scales.